



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/472,249	12/27/1999	ATSUO SHIMADA	0557-4885-2	8492

22850            7590            08/13/2003

OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
1940 DUKE STREET  
ALEXANDRIA, VA 22314

[REDACTED] EXAMINER

BIENEMAN, CHARLES A

ART UNIT	PAPER NUMBER
2176	[REDACTED]

DATE MAILED: 08/13/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Offic Action Summary</b>	Application No.	Applicant(s)
	09/472,249	SHIMADA ET AL.
	Examiner Charles A. Bieneman	Art Unit 2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) Responsive to communication(s) filed on 27 December 1999 and 27 December 2000.
- 2a) This action is FINAL.                  2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) Claim(s) 1-41 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) \_\_\_\_\_ is/are allowed.
- 6) Claim(s) 1-41 is/are rejected.
- 7) Claim(s) \_\_\_\_\_ is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 27 December 1999 is/are: a) accepted or b) objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) The proposed drawing correction filed on \_\_\_\_\_ is: a) approved b) disapproved by the Examiner.  
 If approved, corrected drawings are required in reply to this Office action.
- 12) The oath or declaration is objected to by the Examiner.

#### Priority under 35 U.S.C. §§ 119 and 120

- 13) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some \* c) None of:
1. Certified copies of the priority documents have been received.
2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_ .
3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.
- 14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
- a) The translation of the foreign language provisional application has been received.
- 15) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413) Paper No(s). _____  |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                               | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449) Paper No(s) <u>4, 5, 7</u> . | 6) <input type="checkbox"/> Other: _____                                    |

## DETAILED ACTION

1. This action is responsive to the following communications: original application filed on December 27, 1999, Information Disclosure Statements filed on December 27, 1999 (paper nos. 4 and 5), December 27, 2000 (paper no. 7), and July 9, 2003 (paper no. 9), and priority papers filed December 27, 1999.
2. Claims 1-41 are pending. Claims 1, 7, 8, 14, 24, 30, 31, and 37-41 are independent claims.

### *Priority*

3. Receipt is acknowledged of papers submitted under 35 U.S.C. 119(a)-(d), which papers have been placed of record in the file.
4. Acknowledgment is made of applicant's claim for priority under 35 U.S.C. 119(a)-(d) based upon an application filed in Japan on December 24, 1999. A claim for priority under 35 U.S.C. 119(a)-(d) cannot be based on said application, since the United States application was filed more than twelve months thereafter. *Claim Objections*

5. **Claim 1** is objected to because of the following informalities: "documents memory" in line 6 should be "document memory" as was recited in line 4. Appropriate correction is required.
6. **Claim 2** is objected to because of the following informalities: there should be an article, *i.e.*, "a", preceding the phrase "totalization unit" in line 5. Appropriate correction is required.
7. **Claims 9-11** are objected to because of the following informalities: there should be an article, *i.e.*, "the" or "said", preceding the phrase "vector creation unit". Appropriate correction is required.

8. **Claim 12** is objected to because of the following informalities: there should be an article, *i.e.*, “a”, preceding the phrase “selection information appending unit” in line 2. Appropriate correction is required.
9. **Claim 14** is objected to because of the following informalities: “the above document data” is recited in lines 18-19. While it is clear that antecedent basis for this limitation lies in the “document data” recited in line 5, the word above, besides being redundant and unnecessary, is confusing. Appropriate correction is required.

***Claim Rejections - 35 USC § 102***

10. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

11. With respect to the rejection of each dependent claim below, the preceding rejection(s) of the relevant base claim(s) is incorporated therein.
12. **Claims 1, 4-6, 24, 27-29, and 38** are rejected under 35 U.S.C. 102(b) as being anticipated by Patent Abstract of Japan Publication Number 0711452, published May 2, 1995 (hereinafter “Sharp Corp.”), provided by applicants in their Information Disclosure Statement.

Regarding **independent claims 1, 24, and 38**, Sharp Corp. teaches a document memory which stores input document data. (Sharp Corp., Constitution, lines 1-2: “storage part 101 where document data is stored”.)

Further, Sharp Corp. teaches a selection unit which selects all or part pf document data stored in the document memory. (Sharp Corp., Constitution, lines 2-3: “document analysis part 102 which analyzes document data”.)

Further, Sharp Corp. teaches a characteristics extraction unit which extracts data relating to characteristics of letter rows extracted by the characteristics extraction unit. (Sharp Corp., Constitution, lines 3-5: “word vector generating part 103 . . .”)

Further, Sharp Corp. teaches a work processing unit which work-processes all or part of the document data based on the data relating to characteristics of letter rows extracted by the characteristics extraction unit. (Sharp Corp., Constitution, lines 10-12: “classifying part 107 . . .”)

Further, Sharp Corp. teaches an output unit which outputs all or part of the document data work-processed by the work processing unit. (Sharp Corp., Constitution, lines 12-13: “result storage part . . .”)

Regarding **dependent claims 4 and 27**, Sharp Corp. teaches that the document memory further stores all or part of the document data work-processed by the work processing unit. (Sharp Corp., Constitution, lines 12-13: “result storage part . . .”)

Regarding **dependent claims 5 and 28**, Sharp Corp. teaches the selection unit further selecting all or part of the document data output by the output unit. (Sharp Corp., Constitution, lines 13-15: inherent in “feature vector generating dictionary . . .”)

Regarding **dependent claims 6 and 29**, Sharp Corp. teaches the document memory further storing data relating to the contents of the work processing. (Sharp Corp., Constitution, lines 13-15: “feature vector generating dictionary . . .”)

***Claim Rejections - 35 USC § 103***

13. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

14. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

15. **Claims 2-3 and 25-26** are rejected under 35 U.S.C. 103(a) as being unpatentable over Sharp Corp. in view of U.S. Patent Number 5,880,742 to Rao et al., issued March 9, 1999, filed November 14, 1996, provided by applicants in their Information Disclosure Statement filed December 27, 2000.

Regarding **dependent claims 2 and 25**, Sharp Corp. teaches that the output unit comprises an item value set unit which set a plurality of item values based on the contents of all or part of the document data work-processed by the work-processing unit. (Sharp Corp., Constitution, lines 7-9: “document vector generating part 105 . . .”)

Further, Sharp Corp. does not teach a totalization unit which totalizes all or part of the document data for each item value set by the item value set unit or that the output unit outputs all

or part of the document data in the form of a table having an item value as at least one axis.

However, Rao et al. teach such a totalization unit inasmuch as they teach totalizing item values.

(Rao et al., col. 6, lines 52-58: “presenting an image of a spreadsheet having a plurality of cells arranged in a plurality of rows and columns, each of the cells including an image of character (non-graphical) information representing a respectively paired one of a plurality of data items stored in a data array in the memory of the system, each data item indicating character (non-graphical) information.”) Moreover, one of ordinary skill in the art would have recognized that the user would have gained benefit from being able to see item value data totalized and presented in an easy to read table format. Therefore, it would have been obvious to one of ordinary skill in the art to implement a totalization unit which totalizes all or part of the document data for each item value set by the item value set unit and an output unit outputting all or part of the document data in the form of a table having an item value as at least one axis.

Regarding **dependent claims 3 and 26**, Sharp Corp. does not teach outputting all or part of the document data work-processed by the work processing unit together with all or part of the document data in its state prior to work processing by the work processing unit. However, Rao et al. teach displaying work-processed document data alongside such data in the state it was in prior to being work-processed. (Rao et al., col. 24, lines 59-63: “Two cell regions 26 and 27 are marked in FIG. 14 as examples of the direct symbolic representation of data in the cell region 27 of the focal region and the indirect, graphical representation of data in the cell region 26 of the context area.”) Moreover, one of ordinary skill in the art would have recognized that users would want to be able to see the underlying data being represented by the work-processed data in order to better understand what they were looking at when they saw the work-processed data.

Therefore, it would have been obvious to one of ordinary skill in the art to implement outputting all or part of the document data work-processed by the work processing unit together with all or part of the document data in its state prior to work processing by the work processing unit.

16. **Claims 7-13, 30-36, and 39-40** are rejected under 35 U.S.C. 103(a) as being unpatentable over Sharp Corp. in view of Douglas R. Cutting et al., "Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections", *15<sup>th</sup> Annual International SIGIR* (ACM: 1992), pages 1-12, provided by applicants in their Information Disclosure Statement filed December 27, 1999.

Regarding **independent claims 7, 30, and 39**, Sharp Corp. teaches an input unit which inputs document data. (Sharp Corp., Constitution, lines 1-2: inherent in "storage part 101 . . .")

Further, Sharp Corp. teaches a language analyzer unit which analyzes document data and obtains language analysis information. (Sharp Corp., Constitution, lines 2-3: "document analysis part . . ."; Purpose, line 1: "To use semantic differences . . .")

Further, Sharp Corp. teaches a vector creation unit which obtains document characteristic vectors for the document data based on the language analysis information. (Sharp Corp., Constitution, lines 6-8: "document vector generating part . . .")

Further, Sharp Corp. teaches a classification unit which classifies documents based on the degree of similarity between document characteristic vectors created by the vector creation unit. (Sharp Corp., Constitution, lines 10-13: "classifying part . . .")

Further, Sharp Corp. does not teach creating clusters of documents. However, Cutting et al. teach that document clustering was well known in the art to improve document search and

retrieval. (Cutting et al., page 1, col. 1, lines 18-20.) Therefore, it would have been obvious to one of ordinary skill in the art to create clusters of document.

Further, Sharp Corp. does not teach a cluster characteristics calculation unit. However, such a unit is inherent in Cutting et al.'s teaching of presenting summaries of clusters. (Cutting et al., page 2, col. 2, lines 6-8.) Moreover, one of ordinary skill in the art would have recognized that cluster characteristics would have been relevant to the user because the user would have gained the advantages of clusters by being able to use their characteristics, i.e., whatever attribute(s) the clustered documents had in common. Therefore, it would have been obvious to one of ordinary skill in the art to implement a cluster characteristics calculation unit.

Further, Sharp Corp. teaches a classification category memory (Sharp Corp., Constitution, lines 12-13: "result storage part . . ."), but does not teach that this memory stores cluster characteristics. However, in view of the obviousness of calculating cluster characteristics noted above, it further would have been obvious to one of ordinary skill in the art to implement a cluster classification memory storing cluster characteristics so that the cluster characteristics could have been used and reused in future search and retrieval efforts.

Regarding **independent claims 8, 31, and 40**, the rejection of claim 7 above is fully incorporated herein.

Further, Sharp Corp. does not teach a display unit which displays cluster characteristics calculated by the cluster characteristics calculation unit. However, Cutting et al. inherent teach such a display unit inasmuch as they teach displaying calculated cluster characteristics to the user. (Cutting et al., page 2, col. 2, lines 6-8.) Moreover, one of ordinary skill in the art would have recognized that the user would have needed to see the cluster characteristics in order to

know which ones to select, and therefore the recited display unit would have been obvious to one of ordinary skill in the art.

Further, Sharp Corp. do not teach a cluster selection specification unit which selects predetermined clusters from clusters of documents created by the classification unit. However, such a cluster selection specification unit is inherent in Cutting et al.'s teaching of selecting clusters. (Cutting et al., page 2, col. 2, lines 8-10.) Moreover, one of ordinary skill in the art would have recognized that the user would need to select clusters in order to refine a search or to get search results. Therefore, it would have been obvious to one of ordinary skill in the art to implement the recited cluster selection specification unit.

Regarding **dependent claims 9 and 32**, Sharp Corp. teaches document characteristic vector memory which stores characteristic vectors created by the vector creation unit. (Sharp Corp., Constitution, lines 9-10: "document vector storage part . . .")

Further, Sharp Corp. does not teach a vector correction unit which corrects document vector characteristics in the document characteristic vector memory, so that document characteristic vectors of documents belonging to clusters selected by the cluster selection unit are deleted. However, such a vector correction unit is suggested by Cutting et al.'s teaching on page 2, col. 2, lines 4-17, of removing documents from a cluster as the definition of the cluster is narrowed down. Therefore, it would have been obvious to one of ordinary skill in the art to implement a vector correction unit which corrects document vector characteristics in the document characteristic vector memory, so that document characteristic vectors of documents belonging to clusters selected by the cluster selection unit are deleted.

Further, Sharp Corp. teaches a classification unit as noted above regarding claim 7, but does not teach that documents are classified based on the document characteristic vectors corrected by the vector correction unit. However, this step is suggested by Cutting et al. under the rationale stated in the preceding paragraph.

Regarding **dependent claims 10-11, and 33-34**, Sharp Corp. teaches document characteristic vector memory which stores characteristic vectors created by the vector creation unit. (Sharp Corp., Constitution, lines 9-10: “document vector storage part . . .”)

Further, Sharp Corp. does not teach a document expression space correction unit which corrects document expression space when determining the degree of similarity between document characteristic vectors stored in the document characteristic vectors memory, based on a characteristics amount calculated from the clusters selected by the cluster selection unit. However, such a document expression space unit is suggested by Cutting et al.’s teaching on page 2, col. 2, lines 4-17, of removing documents from a cluster as the definition of the cluster is narrowed down. Therefore, it would have been obvious to one of ordinary skill in the art to implement a document expression space correction unit which corrects document expression space when determining the degree of similarity between document characteristic vectors stored in the document characteristic vectors memory, based on a characteristics amount calculated from the clusters selected by the cluster selection unit.

Further, Sharp Corp. teaches a classification unit as noted above regarding claim 7, but does not teach that documents are classified based on the degree of similarity between document characteristic vectors created by the vector creation unit. However, this step is suggested by Cutting et al. under the rationale stated in the preceding paragraph.

Regarding **dependent claims 12 and 35**, Sharp Corp. do not teach a selection information appending unit which appends selection information showing the fact of selection when all or part of the documents belong to a cluster have been selected. However, Cutting et al. inherently teach such a selection information appending unit inasmuch as they teach gathering selected groups together to form a subcollection. (Cutting et al., page 2, col. 2, lines 10-11.) Moreover, one of ordinary skill in the art would have recognized that the fact that a document had been selected was relevant to the fact of whether it belonged to the cluster and was of interest to the user. Therefore, it would have been obvious to one of ordinary skill in the art to implement a selection information appending unit which appends selection information showing the fact of selection when all or part of the documents belong to a cluster have been selected.

Further, Sharp Corp. do not teach that the display unit displays cluster characteristics, but this limitation would have been obvious in view of Cutting et al. as discussed above regarding claim 8. Further, display of selection information is inherent in Cutting et al.'s teaching of successive iterations of clusters (Cutting et al., page 2, col. 2, lines 4-17); in the process taught by Cutting et al. selection information would have had to have been displayed in order for the user to select for succeeding iterations.

Regarding **dependent claims 13 and 36**, Sharp Corp. does not teach that the classification memory stores cluster characteristics and/or information created by an operator, in addition to all or part of the documents belonging to a cluster of documents selected by the selection specification unit as constituent elements of classification categories. However, Cutting et al. teach storing information created by an operator inasmuch as they teach storing operator selections of groups. (Cutting et al., page 2, col. 2, lines 8-10.) Moreover, one of

ordinary skill in the art would have recognized that it would have been relevant to the clustering process to have known what the user selected. Therefore, it would have been obvious to one of ordinary skill in the art to that the classification memory stores cluster characteristics and/or information created by an operator, in addition to all or part of the documents belonging to a cluster of documents selected by the selection specification unit as constituent elements of classification categories.

17. **Claims 14-16, 22-23, 37, and 41** are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent Number 6,192,360 B1 to Dumais et al., issued February 20, 2001, filed June 23, 1998, in view of Rao et al.

Regarding **independent claims 14 and 37**, Dumais et al. teach a document input unit which inputs document data groups inasmuch as Dumais et al. teach providing document training data including category labels. (Dumais et al., col. 9, lines 21-23.)

Further, Dumais et al. teach a document dividing unit which divides document data into one or more multiple divided document data based on a predetermined reference inasmuch as Dumais et al. teach subjecting documents to a feature extraction process. (Dumais et al., col. 22, lines 42-56.)

Further, Dumais et al. do not teach a document-divided map creation unit which creates a map showing the correspondence between the document data and the divided document data. However, Rao et al. teach mapping data in a spreadsheet, the equivalent of document data, with categories into which the data has been divided. (Rao et al., col. 24, lines 59-63: "Two cell regions 26 and 27 are marked in FIG. 14 as examples of the direct symbolic representation of data in the cell region 27 of the focal region and the indirect, graphical representation of data in

the cell region 26 of the context area.”) Moreover, one of ordinary skill in the art would have recognized that the user would have wanted to see how the document data had been divided. Therefore, it would have been obvious to one of ordinary skill in the art to implement a document-divided map creation unit which creates a map showing the correspondence between the document data and the divided document data.

Further, Dumais et al. teach a divided document classification unit which classifies the divided document data and a divided document classification result creation unit which creates divided document classification result information based on a classification result of the divided document classification unit. (Dumais et al., col. 22, lines 57-60: “The reduced feature vector generated by the feature reduction process 220 (or the binarized reduced feature vector generated by the binarizing process 230) is applied to one or more classifiers 250.”)

Further, Dumais et al. teach a document classification result creation unit which creates classification results information of the above document data using the document-divided document map and the divided document classification result information. (Dumais et al., col. 22, lines 65-67: “A probability that the textual information object 205 belongs to a particular class is output.”)

Regarding **dependent claim 15**, Dumais et al. teach document save unit which saves the document data inasmuch as the document data would have had to have been saved before it could have been used in block 205 of Figure 2.

Further, Dumais et al. teach a divided document save unit which saves the divided document data inasmuch as the reduced feature vectors taught by Dumais et al. would have had to have been saved before they could be used in the classification step in Figure 2.

Further, Dumais et al. do not teach a document-divided map save unit which saves a document-divided document map created by the document-divided document map creation unit. However, such a document-divided map save unit would have been obvious to one of ordinary skill in the art in view of Rao et al. under the rationale discussed above regarding claim 14 and because one of ordinary skill in the art would have recognized that it would have been desirable to save the map for the user's future use.

Regarding **dependent claim 16**, Dumais et al. teaches a divided document classification result save unit which saves the divided document classification result information created by the divided document classification result creation unit inasmuch as saving classification results would have been necessary when classifying training documents as taught by Dumais et al. in order to use those results later. (Dumais et al., col. 9, lines 19-21.)

Regarding **dependent claims 22 and 23**, Dumais et al. teach extracting and presenting information showing document data and divided document data, and representative information accompanying the document data and divided document data, as classification result information. (Dumais et al., Fig. 9; col. 13, lines 22-30.)

18. **Claims 17-18 and 20** are rejected under 35 U.S.C. 103(a) as being unpatentable over Dumais et al. in view of Rao et al. and further in view of U.S. Patent Number 6,128,410 to Park et al, issued October 3, 2000, filed June 8, 1998.

Regarding **dependent claim 17**, Dumais et al. does not teach a plurality of divided document data comprising the document data in its state prior to being divided. However, Park et al. teach divided document data comprising the document data in its state prior to being divided. (Park et al., col. 3, lines 39-43: "The feature extractor 102 extracts feature values of the

input character, representing its original characteristics, from the pre-processed input pattern, e.g., the number of black pixels within a partial area or direction of strokes, to make feature vectors (step 300).") Moreover, one of ordinary skill in the art would have recognized that document data in the state it was in prior to being divided would have more accurately represented the document. Therefore, it would have been obvious to one of ordinary skill in the art to have a plurality of divided document data comprising the document data in its state prior to being divided.

Regarding **dependent claim 18**, Dumais et al. does not teach dividing document data based on information relating to the structure of the document data. However, Park et al. teach divided document data based on patterns equivalent to structure. (Park et al., col. 3, lines 33-34.) Moreover, one of ordinary skill in the art would have recognized that document structure was one way to determine the category to which a document belonged. Therefore, it would have been obvious to one of ordinary skill in the art to divide document data based on information relating to the structure of the document data.

Regarding **dependent claim 20**, Dumais et al. does not teach dividing the document data in compliance with a specified specification range. However, Park et al. teach dividing a pattern equivalent to document data according to the distance of pixels from an arbitrary model M, *i.e.*, a specified specification range. (Park et al., col. 3, lines 44-48.) Moreover, one of ordinary skill in the art would have recognized that it would be disadvantageous to infinitely divide document data and that therefore it would be advantageous to divide the document data in compliance with a specified specification range. Therefore, it would have been obvious to one of ordinary skill in the art to divide the document data in compliance with a specified specification range.

19. **Claim 19** is rejected under 35 U.S.C. 103(a) as being unpatentable over Dumais et al. in view of Rao et al. and further in view of Sharp Corp.

Regarding **dependent claim 19**, Dumais et al. do not teach a document element extraction unit which extracts elements in the document data or an element-accompanying information extraction unit which extracts element-accompanying information accompanying the elements extracted by the element accompanying information extraction unit. However, Sharp Corp. teaches these limitations inasmuch as Sharp Corp. teaches extracting word vectors and feature vectors. Moreover, one of ordinary skill in the art would have recognized that word were a basic document element and occurrence of words in a document would have been relevant to classifying the document. Therefore, it would have been obvious to one of ordinary skill in the art to implement a document element extraction unit which extracts elements in the document data and an element-accompanying information extraction unit which extracts element-accompanying information accompanying the elements extracted by the element accompanying information extraction unit.

Further, Dumais et al. do not but Sharp Corp. does teach dividing the document data using elements extracted by the document element extraction unit, or the elements and element-accompanying information extraction unit. (Sharp Corp., Constitution, lines 6-9: “document vector generating part . . . generates feature vectors of the document from feature vectors of words included in the document . . .”) Moreover, one of ordinary skill in the art would have recognized that it would have been logical to divide document data according to basic elements in the document such as words. Therefore, it would have been obvious to one of ordinary skill in

Art Unit: 2176

the art to divide the document data using elements extracted by the document element extraction unit, or the elements and element-accompanying information extraction unit.

20. **Claim 21** is rejected under 35 U.S.C. 103(a) as being unpatentable over Dumais et al. in view of Rao et al. and further in view of U.S. Patent Number 6,581,056 B1 to Rao et al., issued June 17, 2003, filed June 27, 1996 (hereinafter “Rao II”).

Regarding **dependent claim 21**, Dumais et al. do not teach dividing document data based on the number of letters, the number of sentences, or both. However, Rao II teach dividing document data based on the statistical occurrences of token representing character strings. (Rao II, col. 5, lines 60-63.) Notice is taken that it was well known in the art that words and sentences were comprised of character strings. Moreover, one of ordinary skill in the art would have recognized that the number of times a word or sentence occurred would have been relevant to classifying the document in which the word or sentence occurred. Therefore, it would have been obvious to one of ordinary skill in the art to divide document data based on the number of letters, the number of sentences, or both.

### *Conclusion*

21. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

U.S. Patent Number	Name	Issue Date	File Date
6,562,077 B2	Bobrow et al.	5/13/03	11/14/97
6,542,635 B1	Hu et al.	4/1/03	9/8/99
6,463,426 B1	Lipson et al.	10/8/02	10/26/98
6,249,779 B1	Hitt	1/19/01	5/13/98
6,212,532 B1	Johnson et al.	4/3/01	10/22/98
6,125,362	Elworthy	9/26/00	12/4/97
5,893,092	Driscoll	4/6/99	6/23/97
5,899,992	Iyer et al.	5/4/99	2/14/97
5,717,914	Husick et al.	2/10/98	9/15/95

Art Unit: 2176

5,710,916	Barbara et al.	1/20/98	1/16/95	
5,454,104	Steidlmaier et al.	9/26/95	n/a	
5,072,367	Clayton et al.	12/10/91	n/a	

22. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Charles A. Bieneman whose telephone number is 703-305-8045. The examiner can normally be reached on Monday - Thursday, 6:30 a.m. - 5:00 p.m..

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph H. Feild can be reached on 703-305-9792. The fax phone numbers for the organization where this application or proceeding is assigned are 703-746-7239 for regular communications and 703-746-7238 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-4700.

CAB  
July 30, 2003



SANJIV SHAH  
PRIMARY EXAMINER